Parameter Estimation: Some General Principles

Population Modeling
University of Florida
Gainesville, FL
February-March 2016

USGS
science for a changing world

Motivation I

- Course focuses on matrix population models
- These models are populated by so-called "vital rates":
  - Survival probabilities
  - Reproductive rates
- Require good estimates of vital rates

Motivation II

- Useful to directly estimate λ using
  - Capture-recapture data
  - Abundance survey estimates
  - Integrated population models
- Useful to directly estimate contributions of demographic components to λ

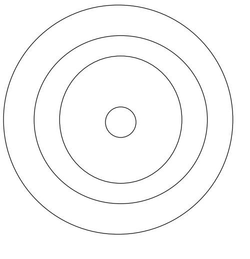
Some definitions

- Parameter
  - Unknown, (fixed) quantity associated with population
- Statistic
  - Summary of sample measurements
  - If sample taken randomly, then also random variable
- Estimator
  - Statistic associated with a parameter
  - Denoted by ^ ("hat")
- Estimate
  - Specific number from an estimator
  - Vary from sample to sample

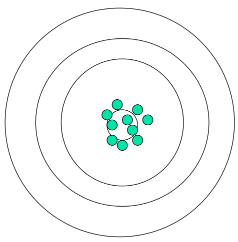
How do estimators behave?

- Probability density (mass) function
  - Frequency distribution of an estimator over multiple samples
    - Plot of f(x) vs x
    - E.g. coin flips: x: {0,1}, f(x): 0.5, 0.5
    - Represents frequency from population
    - Frequencies sum (or integrate) to 1
    - Non-negative
    - Discrete (coin toss)
    - Continuous (means)
- Describes behavior of random variables
  - Assigns probabilities to outcomes

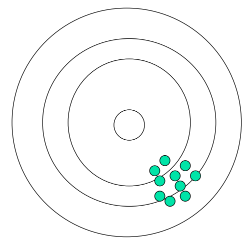
Concepts: Bias and Precision



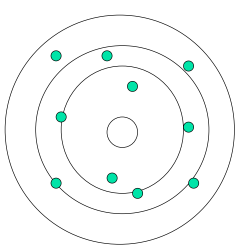
Unbiased and Precise



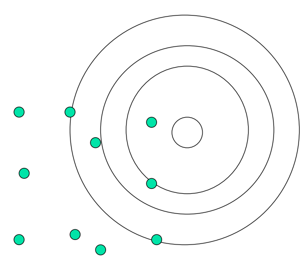
Biased but Precise



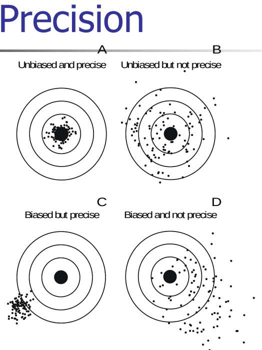
Unbiased but Imprecise



Biased and Imprecise



Bias, Precision



Attributes of Estimators

- Expected value $\mathbf{E}(Y) = \Sigma Y * f(y)$
  - Weighted average of R.V. over all possible values
  - Mean value of die roll:

$$E(x) = 1*(1/6) + 2*(1/6) + ... + 6*(1/6) = 3.5.$$

- Bias: $\mathbf{E}(\hat{Y}) - Y$
  - Difference between expected value of estimate and parameter (MLE)
- Variance: $var(\hat{Y}) = \mathbf{E}([\hat{Y} - \mathbf{E}(\hat{Y})]^2)$
- Mean Squared Error
  - $\mathbf{E}([\hat{Y} - Y]^2) = Var(\hat{Y}) + [Bias(\hat{Y})]^2$

Accuracy
(balance of bias and precision)

Mean Square Error

$$MSE(\hat{S}) = Variance + Bias^2$$

Concepts and Notations

Pr(**X**|**θ**) – probability of observing data **X** given population parameters **θ**

Concepts and Notations

- Methods of Inference
  - Maximum Likelihood
  - Bayesian

Example: a coin tossing experiment

Suppose a coin is tossed 5 times with the result:

X = HHTHT

Assuming each toss is independent,

$$\Pr(\text{HHTHT} \mid p) = pp(1-p)\,p(1-p)$$
$$= p^3(1-p)^2$$

Concepts and Notations

Probability statements could be used to estimate parameters using either
maximum likelihood
or
Bayesian
methods of inference.

Likelihood

- For discrete data, a 'likelihood' is simply the probability of observing the data.
- In a probability statement, the data are conditional upon the parameters.
- This is reversed in the likelihood, e.g.,

  L(θ | X) = Pr(X | θ)
- Find what parameter values maximize the likelihood function (i.e., MLE).

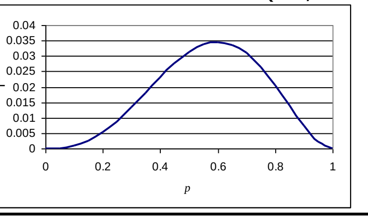
Likelihood

- The degree of curvature in the likelihood function about the MLE reflects the estimator variance.
- Technically, variances are obtained from the matrix of second partial derivatives of the likelihood function.

Example: a coin tossing experiment

Likelihood: L(p|HHTHT) = $p^3(1-p)^2$

- find the value of p that maximizes the likelihood of observing this sequence of 3 heads from 5 coin tosses (i.e., the MLE).



Bayesian

Application of Bayes Theorem

$$\Pr(\boldsymbol{\theta}|\mathbf{X}) = \frac{\Pr(\boldsymbol{\theta})\Pr(\mathbf{X}|\boldsymbol{\theta})}{\Pr(\mathbf{X})}$$

posterior distribution of **θ**

generally unknown, but not required with Markov chain Monte Carlo approaches.

prior distribution of **θ**

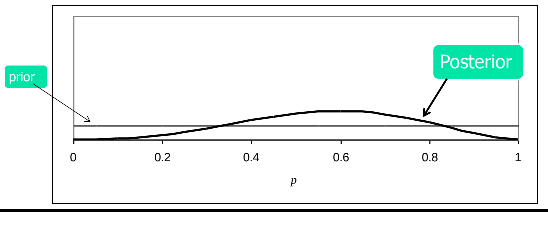
Bayesian

Resulting inferences may be sensitive to choice of prior distribution, particularly when the data contain little information about the parameter(s).

Example: a coin tossing experiment

Bayesian:

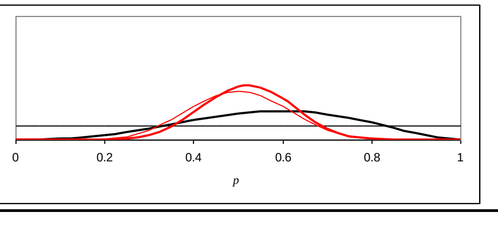
- Use the data to update prior 'knowledge' of the distn. to obtain the posterior distn. of p



Example: a coin tossing experiment

Bayesian:

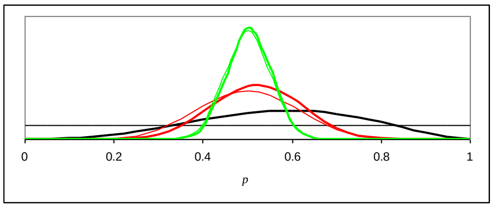
- Use the data to update prior 'knowledge' of the distn. to obtain the posterior distn. of p



Example: a coin tossing experiment

Bayesian:

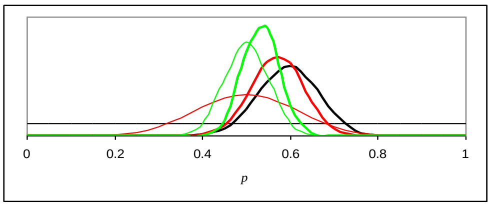
- Use the data to update prior 'knowledge' of the distn. to obtain the posterior distn. of p



Example: a coin tossing experiment

Bayesian:

- Data with high information content can overwhelm stronger priors, e.g., 30 heads from 50 coin tosses



Max. Likelihood vs. Bayesian

	ML	Bayesian
Pros	Widely used. Computationally straightforward.	More natural interpretation. Easier to implement complex models. Can incorporate prior knowledge.
Cons	Numerical issues (has 'true' maximum been found?). 'Unnatural' interpretation.	How should prior knowledge be expressed. Computationally intensive. Difficult to spot unestimatible parameters.

Incorporating Covariates

- Often we may wish to allow parameters to be functions of covariates or other factors of interest (e.g., habitat type, rainfall in the last 24 hours, distance from a road).
- Generally requires a *link function* to transform linear relationships on to the 0-1 interval.

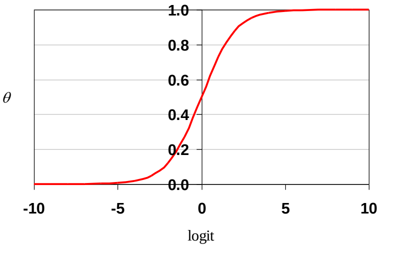
The Logit Link

$$\text{logit}(\theta_i) = \ln\left(\frac{\theta_i}{1-\theta_i}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots$$

which can be rearranged as,

$$\theta_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots)}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots)}$$

The Logit Link



Hypotheses, Theories and Models

- A *hypothesis* is simply a story about how the world (or a part of it) works.
- A *theory* is a hypothesis that has become widely accepted after surviving repeated attempts to falsify it.

Hypotheses, Theories and Models

- A *model* is an abstraction of a system, and may include hypotheses and theories about the system.
- Models are used to describe and predict system behavior.
- A model may be conceptual, verbal or mathematical.

Hypotheses, Theories and Models

- Mathematical models are often developed to project the consequences of hypotheses.
- Competing hypotheses can be represented as different models.
- Model-based prediction of system behavior under competing hypotheses represents a key step in the conduct of science and management.

Comparing models

- Hypothesis testing
  - seek 'sufficient' evidence to falsify a null hypothesis
  - e.g., likelihood ratio tests
- Information theoretic approaches
  - rank models in order of relative distance from 'truth'
  - model weights can be calculated
  - e.g., AIC, DIC, etc
- Bayesian approaches
  - Bayes factors
  - directly estimate model probabilities given the data

Likelihood Ratio Tests

1. Fit 2 models to the data, with and without the effect of interest (θ).
2. Calculate test statistic as
   $$LR = 2\left[l(\theta) - l(\theta')\right]$$
3. Compare to chi-square distribution with degrees of freedom equal to the number of parameters required to estimate θ.

Information Theoretic Methods

- Relative measure of distance from 'truth' based on Kullback-Liebler Information.

$$AIC = -2l(\theta) + 2K$$

- K = number of parameters
- Models with small values are preferred
- Parsimony is useful by-product

Information Theoretic Methods

- Small sample adjustment could be used.

$$AICc = AIC + \frac{2K(K+1)}{n-K-1}$$

- Debate over 'effective' sample size in occupancy models.

Information Theoretic Methods

- Can adjust AIC for overdispersion/poor model fit.

$$QAIC = \frac{-2l(\theta)}{\hat{c}} + 2K$$

$$= \frac{1}{\hat{c}}\left[AIC + 2(\hat{c}-1)K\right]$$

$$\propto AIC + 2(\hat{c}-1)K$$

Information Theoretic Methods

- ΔAIC = AIC –min(AIC) is what is important.

Information Theoretic Methods

- "Likelihood" of model *j*, given the data.

$$L(\text{model } j|\text{data}) = \exp(-\Delta IC_j/2)$$

Information Theoretic Methods

- Model weights can be calculated that are ad-hoc measures of support for each model in the candidate set

$$w_j = \frac{\exp(-\Delta IC_j/2)}{\sum_{l=1}^{m}\exp(-\Delta IC_l/2)}$$

- Can be used to obtain model averaged estimates, or summed across models to evaluate the importance of specific factors.

Information Theoretic Methods

- Evidence ratios

$$E_{jk} = \frac{w_j}{w_k}$$

- A measure for the degree of support for model *j* compared to model *k*.

Model Averaging

- May want to make inferences from multiple models.

$$\hat{\theta}_A = \sum_{l=1}^{m} w_l \hat{\theta}_l$$

$$SE\left(\hat{\theta}_A\right) = \sum_{l=1}^{m} w_l \sqrt{Var\left(\hat{\theta}_l \mid M_l\right) + \left(\hat{\theta}_l - \hat{\theta}_A\right)^2}$$

43

Model Averaging

- Care must be taken when model averaging β parameters that respective parameters in different models have the same interpretation.
- Estimated β parameters can be sensitive to other covariates included in model (e.g., interactions or correlations).
- Suggest model averaging should be performed on the focal parameters (i.e., the survival or detection probabilities).

44

"Absolute" fit of model

- "A third principle recommends thorough checks on the fit of a model to the data … Such diagnostic procedures are not yet formalized, and perhaps never will be. Some imagination or introspection is required…to determine the aspects of the model that are most important and most suspect." (McCullagh & Nelder 1989, Chapman and Hall, p. 8)

Pearson's χ^2 GOF Test

- Logic: If model is 'correct', expected and observed cell frequencies for each multinomial cell should be similar.
- If sample size is adequate, *(expect at least 2 per cell),*

$$\Sigma(\text{observed}_i - \text{expected}_i)^2/\text{expected}_i$$

$$\sim \chi^2(\text{df} = \#\text{ cells} - 1)$$

USGS

Bootstrap GOF Test

- Compute ML estimates for parameters,
- Produce empirical distribution of estimates:
  - Simulate capture histories for each released animal:
    - assume parameter = MLE,
    - 'flip coins' to determine survival and capture for each period,
  - Repeat for $\{\boldsymbol{R}_i\}$ animals, estimate parameters,
  - Compute deviance
- Compare original deviance with empirical distribution (i.e., what percentile?)